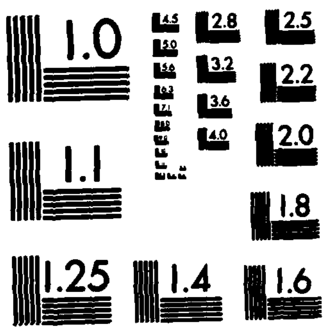
AD-A144 438 MATHEMATICAL MODELS OF THE EVENT RELATED POTENTIAL(U) 1/1
WASHINGTON UNIV SEATTLE DEPT OF PSYCHOLOGY E B HUNT
01 APR 84 AFOSR-TR-84-0699 AFOSR-83-0289

UNCLASSIFIED F/G 6/16 NL

END



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AFOSR-TR- 84-0699

AD-A144 438

MATHEMATICAL MODELS OF THE EVENT RELATED POTENTIAL

EARL HUNT

APRIL 1984

This research was sponsored by:

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

Project-Task 2313/A4

Grant No. AFOSR-83-0289

84 08 17 064

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR-TR- [illegible]	2. GOVT ACCESSION NO. AD-A144438	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) MATHEMATICAL MODELS OF THE EVENT RELATED POTENTIAL		5. TYPE OF REPORT & PERIOD COVERED INTERIM 83 Sep 30 - 84 April 1
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Earl B. Hunt		8. CONTRACT OR GRANT NUMBER(s) AFOSR-83-0289
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Psychology, University of Washington, Seattle, Washington 98195		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2313/A4
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research (NL) Bolling AFB DC 20332		12. REPORT DATE April 1, 1984
		13. NUMBER OF PAGES 26
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)
Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)
Event-related-potential, electrophysiology, wave forms, principal component analysis, factor analysis.

In electrophysiology,

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)
The Event Related Potential is assumed to be composed of several underlying component wave forms. Principal Component Analysis is a statistical technique that has been used to uncover the components by analysis of the observed wave form. The mathematical assumptions behind Principal Component Analysis are examined, and their plausibility is questioned. It is pointed out that under certain conditions the component forms may not accurately be recovered by Principal Component Analysis. Under other circumstances violations of some (cont)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Block 20 continued

of the mathematical assumptions does not appear to affect the accuracy of recovery of component waveforms. The points made are illustrated by an analysis of simulated wave forms constructed from known components.

Mathematical Models of the Event Related Potential

Earl Hunt
The University of Washington

The event related potential (ERP) is an electrical signal provided by the brain in response to external stimuli. ERPs in humans are recorded from surface electrodes. The resulting amplitude x time waveform is believed to contain some information about the brain's response to the stimulus (the "signal") but that signal is emersed in extraneous information ("noise") about brain events not associated with cognitive or overt responses. A variety of mathematical techniques have been proposed for isolating signal components of the waveform. In an unusually creative paper, Donchin and Heffely (1979) pointed out that there is a mathematical isomorphism between recordings from electrodes and the data obtained from classic personality and intelligence tests, and that this isomorphism could be used to apply a widely used psychometric method of data analysis, Principal Components Analysis (PCA), to the analysis of ERP records. The Donchin and Heffely proposal has been widely adopted.

This paper is a critique of the logic behind PCA.The mathematical assumptions underlying the method will first be reviewed, stressing the plausability of the assumptions as statements about psychophysiology. The critique will emphasize possible distortions of results that could be introduced in situations in which some of the mathematical assumptions are not defensible on psychophysiological grounds.

The Mathematics of the PCA

The recording from a single electrode is most naturally expressed as a variation in amplitude over time,

$$(1) \qquad x[i,t] = f(t),$$

where x[i,t] is the potential of the ith electrode at time t (t = 1...T), and f is a continuous function of time. This will be called the "functional" representation of the ERP. By convention, the stimulus is presented at time t=0. In practice, x[t] is sampled at discrete time intervals, so the signal can be representated as a vector

$$(2) \qquad x[i] = \begin{bmatrix} x[i,0] \\ \cdot \\ \cdot \\ x[i,t] \\ \cdot \\ \cdot \\ x[i,T-1] \end{bmatrix}$$

in T dimensional space. This will be called the "Euclidean" representation of the ERP.

Because there is a one:one correspondence between equation (1), at discrete time points, and equation (2), any data analysis technique that uncovers regularities in the Euclidean representation should be interpretable in terms of a regularity in the more natural functional representation. This fact is the logical basis for Donchin and Heffely's proposal. Principal Components Analysis is a technique for uncovering regularities in a Euclidean representation. Note, though, that the proposal is based on a "backwards" argument, something that is true in the Euclidean representation must be true in the functional representation. But what about argument in the other direction? The proposal is valid to the extent that the translation from the functional to the Euclidean representation retains what is important in the ERP record. To evaluate this issue one must consider how signal and noise components combine to produce the original functional representation.

The ERP is assumed to be produced by the summation of K component wave forms. These are named by their polarity and the approximate times of their peak amplitudes, as in N100, P300, etc. The first non-zero point of a component wave form will be called its latency. This is slightly different from the convention in psychophysiology, which is to define latency by the time of the peak amplitude. The difference does not affect the logic of the argument to be presented.

The natural way to think of a component wave form is as a potential that assumes a non-zero amplitude at time L[K], and follows a fixed time course, with a maximum absolute excursion (amplitude) A[i,k]. The time course of the wave is fixed across all records, but the amplitude may vary from record to record. Thus each wave form could be thought of as having a standard form, that begins at t=0, a maximum absolute amplitude of 1,and taking the value g[k](t) at time t. A standard form can be converted into an actual component form by multiplying the standard form by the amplitude A[i,k] that characterizes the record;

(3) "Value of kth form in record i at time t" = A[i,k] g[k](t).

The component forms sum to produce a "true" record, y[i,t], where

(4) $$y[i,t] = \sum_{K=1}^{K} A[i,k]\, g[k]\,(t).$$

The observed record is derived from the true record by the addition of an error component e[i](t), that has an expected value of zero and some unknown variance.

The model for the observed record becomes

(5) $x[i,t] = \sum_{k=1}^{K} A[i,k]\ g[k](t) + e[i](t).$

It is desirable to assume that the e[i](t)'s are independent over i (records) and t (time points). This is often unrealistic. For instance, most events that would produce a perturbation in the electrical record would extend over several time points. To avoid this problem, the typical solution is to use as a "record" the average waveform recorded over some fairly large number, n (n 100), of trials observed under theoretically identical conditions. The rationale for this is that the true value y[i,t], should remain constant, and e[i](t) should move to its expectation, zero. Writing X and Y for the summed wave forms,

(6) $E(X[i,t]) = Y[i,t] + E(e[i](t))$
$= Y[i,t]$

thus removing the e[i](t)'s from concern.

In the typical ERP study the problem is to extract information about the underlying components from an analysis of observed records, after the averaging described above. Two types of information are of interest; the standard forms (g[k](t)) for each component waves, and the amplitude A[i,k] of each form component in each record. These values are the results of PCA. In order to define a mathematically tractable problem, however, PCA demands some further assumptions about the data. They will now be described.

What Principal Components Analysis Does

In order to understand how PCA works a geometric presentation is useful. Suppose that the model of the ERP that has been presented were exactly true. It would then be possible to plot any ERP record as a point defined by its value on each of the T time points. This is shown for a swarm of ERP records in figure 1. One can imagine a 'best fitting ellipsoid' that would fit this swarm, as shown in the figure. PCA is a mathematical technique for determining the best fitting ellopsoid and finding its axes. The data points will be plotted in N-dimensional space, so the ellipsoid will have N axes. However, the original dimensions (i.e. the time points) will typically have highly correlated values. Therefore, the points will tend to lie in a K dimensional ellipsoid

----

Figure 1 here

----

embedded in the N-dimensional space. A whimsical example may help to illustrate this. A pizza is an "almost" two dimensional ellipsoid lying in a three dimensional space. Computer programs for PCA extract the axes in descending order of their lengths, until it is felt that further extractions of axes would be statistically unreliable. Several criteria have been offered for making this determination. The conventional criterion is 'the eigenvalue of the component is less than one.' This is an arbitrary mathematical determination, somewhat akin in meaning to 'per cent of variance accounted for.' More justifiable, but more complicated, criteria have also been offered. No attempt will be made to defend any of them. In most practical cases they seem to usually lead to the same decision.

Once the ellipsoid has been found, it is possible to plot the projection of any point on each of its axes. This is shown in Figure 2, which plots a single point, i, with reference to two time axes (t=1, 2) and two ellipse axes (k=1, 2). Next, consider the projection of point i on some <u>time</u> axis, t. From elementary analytic geometry, this will be, in the example

(7) $x[i,t] = B[i,1] \cos(\ [1,t]) + B[i,2]\cos(\ [2,t])$

where B[i,k] is the projection of data point i on the kth axis of the ellipse and θ[k,t] is the angle between the kth axis of the ellipsoid and the ith time dimension. B[i,k] is called the factor score of record i on component k, and cosθ[k,t] is the loading of time dimension t on component k.

The general form of (7), for k dimensions and T time points, is

(8) $X[i,t] = \sum_{K=1}^{K} B[i,k]\cos\theta[k,i]$

finally, suppose that K T, i.e. that the extraction of components has been halted at some point less than the full dimensionality of the space. (To illustrate, suppose that only k=1 were to be considered in figure 2.) Then it would be necessary to introduce a correction term, c, for each data point at each time, to account for a data point's movement in 'higher dimension' than those represented in the ellipsoid. Let C[i](t) be the correction to the ith point on the ith time dimension. Then equation (8) expands to

(9) $X[i,t] = \sum_{K=1}^{K} B[i,k] \cos\ [k,t] + C[i](t).$

Equation (9) maps directly onto equation (5), the mathematical model of the ERP, by substituting A[i,k], g[k](t), and e[i](t) for the B, cosθ, and C terms. Therefore it would seem that PCA solves the problem of analyzing the ERP data to find the valves required by the model.

In a sense, it does. The standard form of a component is determined by the loading of the scores involving a factor (i.e. by the vector (cos [k,t]), varying over t), and the amplitude of a component in a particular record is given by the factor score for that record.

However, the solution depends critically upon two assumptions. These are that the distribution of the data, plotted in the T dimensional time space is accurately characterized as an ellipsoid, and that the axes of the ellipsoid correspond to the components used to generate the data. Both these conditions will be met if the following statements are true:

S.1 The dimensions defining each component are orthogonal. That is, there is no correlation between {g[k](t)} and {g[k](t')}, calculated over t.

S.2 The ERPs are defined as in equation (5). This means that the standard forms {g[k](t)} are invariant over all records.

S.3 The amplitudes { A[i,k] } of the kth component in each of the N ERPs are established by taking N independent samples from a normal distribution with expectation E(A[k]) and variance V(A[k]).

S.4 The error terms e[i](t) are determined by independently sampling from a population with expectation E(e(t)) = 0 and variance V(e(t)).

Figure 2 here

Each of the assumptions can be questioned. The effect of relaxing them will be considered in the next section. Before proceeding to the critique, though, one more issue concerning the normal use of PCA needs to be considered. The wave forms recovered directly from PCA tend to be shallow, and may contain both positive and negative components, even when rather sharply defined, single peaked components appear in the graphic record. Just why this should happen is not clear. One possibility is that the data points are being forced into an elliptical form by the analytic procedure, even though the best fitting ellipse is a poor description of the data. (An example of how this might occur is discussed below). A second possibility is that the data does have a generally elliptical shape, but that the estimation procedure used by PCA does not well recover that shape.

A procedure known as "varimaxing" is used to produce more sharply defined wave forms. What varimaxing does is to rotate the discovered axes so that they are approximately parallel to a few time axes and, more importantly, so that they have zero loadings on (i.e. are orthogonal to) many time axes. Such a procedure avoids shallow waveforms. To see why, consider the mathematical meaning of the "shallow" waves, recovered by PCA. In a shallow wave the coeffecients of the standard form, the values in the set { g[k](t) }, are approximately equal in size. As a result, the variance of these numbers, V(g[k]), will be small. This contrasts with the numbers { g[k](t) } derived from a sharply peaked function, which will have a higher variance. The varimax procedure rotates the component axes to maximize the sum of the variances across all components, i.e. it maximizes

$$V = \sum_{k=1}^{K} V(g[k]). \tag{10}$$

The rotation procedure, however, retains the requirement that the component axes be orthogonal. Subject to this constraint, varimaxing will always produce a sharp function. It does so by using a mathematical criterion for evaluating a function (waveform). The mathematical criterion used does not have a clearcut physiological interpretation.

Critique

Principal component analysis solves the mathematical problem for which it is defined. Any quarrel with PCA as a method of analysis must focus on the problem statement; are equation (5) and assumptions S.1-S.4 reasonable approximations of the way in which the psychophysiological data was gathered?

Some arguments against the PCA assumptions will now be offered. For ease of reading, each assumption will be restated in abbreviated form prior to commenting upon it.

S.1. Orthogonality The component dimensions are orthogonal.

This assumption is highly suspect. Mathematically, a necessary and sufficient condition for orthogonality of the kth and jth component is that there be no correlation between the amplitudes of the kth and jth component, when the correlation is computed across individuals. The computational methods used by PCA ensure that the discovered components satisfy the orthogonality requirement. But consider the physiological interpretation of component amplitudes. The terms A[i,k] and A[i,j] represent the amplitudes of two different components, recorded from the same electrode, recording from the same brain. Any general process that affects the excursion of electrical activities (not the average potential) ought to affect both components. Naturally this would produce a correlation across records.

If the ERP is produced by correlated components, the use of PCA ought to produce fewer components than are actually present. The discovered components should be mixtures of the true components. Interestingly, varimaxing may help in such situations. If the first component discovered in a PCA analysis represents a compromise between several correlated components, varimaxing ought to move the largest (first extracted) PCA component towards the largest of the underlying true components. However the relation between the true components and the subsequently extracted PCA components is unclear.

S.2 The same standard wave form applies to all ERP records.

This assumption is clearly false for at least some interesting cases. Suppose one's sample is made up of people who vary widely in age. There is a substantial body of literature indicating that there are latency differences in the ERP's recorded from young and elderly brains (Ford and Pfefferbaum, 1980), and possibly other differences in shape as well. All of these would be reflected as changes in standard forms. A similar case could be made if individual records are taken from different points on the skull. If the various ERP components emanate from different points in the brain, one would expect the latencies of their appearances to be influenced by the placement of the recording electrode.

Failures of assumption S.2 may be partly responsible for the production of "flat" component forms by the use of PCA without rotation. Suppose that the ERP record is actually produced by a single sharply peaked component wave that varies in latency over individuals. This is shown in Figure 3, panel a. PCA will average the form over records, producing a flat wave, similar to that shown in Figure 3, panel b. Varimaxing may partially correct for the situation, by moving the discovered wave form toward the form that exists in most of the records. While this is usually desireable, the nature of the averaging process is not clear. First, the averaging interacts with the orthogonality requirement, so that after the first PCA component has been defined it is not clear what is being averaged. Second, and perhaps more important, the resulting amplitudes do not correctly reflect (and in general will underestimate) the amplitude of the component wave form in aberrant individuals. Finally, and perhaps most important, the method hides individual differences in component wave forms. These may be of interest in themselves.

Figure 3 here

S.3 The amplitudes { A[i,k] } are determined by independent sampling from a normal distribution with mean u(k) and variance v(k).

This is a particularly important assumption to consider, for it appears that PCA is often practiced in situations where S.3 is likely to be false.

Suppose that PCA is conducted on a data set consisting of N[1] records drawn from population 1 and N[2] records from population 2. For example, some of the records might be drawn from healthy participants, and other records drawn from alcoholics. This will be called the "population" design for brevity. Suppose further that assumption S.3 is true within each population. In this case the data points from each population should trace out their own ellipsoid forms in the Euclidean time representation. This is shown by the solid lines in Figure 4.

Figure 4 here

Under such conditions statement S.3 cannot be true of the data set as a whole, unless the two populations are drawn from identical distributions. If PCA is applied to the entire data set the computer program will define a "best fitting ellipse" for the entire sample that may not be characteristic of either population. An example is shown by the dotted line in Figure 4. In fact, the "component" defined by PCA in a population design is likely to be that linear combination of time points that best discriminates between the two groups. This is a major problem if, as is sometimes done, measures derived from PCA..usually the amplitudes..are used to discriminate beween the groups. (For example, one might perform an analysis of variance on the amplitudes of a component that was defined by an analysis of all the data.) In such a design group identity is the independent variable and the measure derived from the PCA is the dependent variable. Most statistical tests do not apply. because effects associated with the independent variable can have an influence on the definition of the dependent variable.

Occasionaly a study is reported in which the data set consists of N[1] electrode recordings from each of N[2] subjects, for a total of N[1] x N[2] records. This will be called the "individuals" design. One motivation for the individuals design is that N must be greater than T in order to produce any PCA solution, and N should be approximately 5 x T to produce a mathematically stable solution. In an individuals design the ERP records from each person can be regarded as defining a coherent, related subset of the data set. Exactly the same considerations apply to the subsets as applied to the groups in the population design. The data points from ech individual may form their own ellipse, but the PCA program will disregard this. Instead, as in the case of the populations problem, PCA will emphasize components that define differences between individuals, at the expense of ignoring components that are dominant in every person.

Statement S3 may "fail" more subtly. Suppose that the amplitudes of different components co-vary across individuals. Such a situation would arise if there were any individual characteristics that affected the amplitudes of the excursion of all electrical activity in a person's brain. Mathematically, this situation would produce a 'general' factor q[i], such that the amplitude of the kth component in the ith individual would be

(11) A[i,k] = C[k] Q[i] + A'[i,k],

where Q[i] is the ith individual's standardized score on the general component, C[k] is the contribution of the general component to the kth specific component, and A'[i,k] is the individual's value on the specific part of the kth component. Precisely how PCA will analyze this situation will depend upon the values of the coeffecients {C[k]}. The point to note here is that the elliptical form of the data points will be distorted by the existence of the general component.

This effect may interact with distortions introduced by the analysis of two groups as if they were a distinct population, as discussed above.

S.4 The error terms, { e[i](t) }, are independently drawn from normal populations with mean 0 and a variance, v(t), that is characteristic of t.

This assumption is central to PCA. It is almost certain to be false if the records to be analyzed are the ERPs taken on individual trials. A necessary condition for S.4 to be true is that random events that influence a recording at time t be statistically independent of random events influencing the recording at time t'. Furthermore, the events influencing the residuals on one recording, i, should be statistically independent of the events influencing recording j.

This condition is never met at the level of individual ERP records, for any event that caused an error in recording at time t would almost certainly have consequences that would influence recordings in the immediately subsequent time periods. It is trivially easy to mitigate such effects by averaging over trials, and using the averaged ERP as the record to be analyzed. The average process is surprisingly efficient, as can be shown by the investigation of extreme cases. Suppose, for instance that the potentials recorded at times t and t+1 have a "true" correlation, (i.e. correlation due to common underlying components) of .3, and the correlation between error effects is .95. Suppose further that the variance of the error distribution is four times the variance of the underlying signal, i.e. a signal/noise ratio of 1:4. The correlation between single records has an expected value of .84. almost entirely due to the error component. If the records are composed of an average of 100 trials the correlations drops to .33, close to the true value and it drops further to .31 after averaging over 500 trials. Clearly the averaging process can drastically reduce the effects of violations of the statement S.4.

Unfortunately averaging has no effects on statements S1, S2 and S3. Any violation of these statements in ERP records obtained on single trials will also apply to the records obtained by averaging over trials. It is therefore reasonable to ask what the effects of these violations is. The only feasible way to answer this question is to analyze "simulated" records, obtained by summing known components, plus error, into "psuedo-ERP" and then seeing if PCA can extract the original components from the constructed records.

Wood and McCarthy (in press) have reported one such study, that was concerned primarily with violations of statements S.1 (orthogonality of the wave forms) and S.3 (component amplitudes defined by sampling from a single population.) Wood and McCarthy found that in this situation PCA could recover the forms of the component waves quite well, but that it might substantially misallocate the percentage of the variance accounted for by all except the first component extracted by the analysis. In practice, this meant that PCA did not correctly identify the component amplitudes in individual records (the values of {A[i,k]}, k 2) for the later components. Wood and McCarthy further found that statistical analyses of inter-group differences in the mean values defined PCA-discovered components did not always agree with a parallel analysis of the (known) components used to generate the simulated ERP records. The latter result would be expected if the individual component amplitude values were mis-estimated.

My colleagues and I have replicated Wood and McCarthy, using a slight variation of their techniques for generating ERP, and have obtained essentially the same results. In addition, we have conducted an analysis of the effect of correlated component amplitudes. 'Psuedo-ERPs' were generated using the three component wave studied by Wood and McCarthy, but with the addition of a randomly generated "general factor", (Q[i] in equation (11)) to each of the amplitudes (A[i,k], k=1...3). The variance of the Q[i]'s was chosen to induce a correlation of .15 between amplitudes on two of the components, calculated across records (i). Table 1 presents the results of this study. It shows the correlations between the amplitudes of each component, as assigned in the development of the ERP record, and the amplitude recovered by PCA followed by varimaxing. These correlations are unacceptably low. It is worth noting that this result is clearly due to the introduction of a correlation between component amplitudes. A repetition of the simulation using uncorrelated amplitudes produced correlations between the 'true' and 'recovered' amplitudes that were all in excess of .95.

Table 1 Here

This result raises a question about the use of PCA both when multiple records are obtained from one person or when the analysis disregards the fact that subjects fall into distinct groups. If the groups or individuals differ in their amplitudes of more than one component, then a correlation between component amplitudes will automatically be introduced by the experimental design (Figure 4). Thus the practice of applying PCA to all the records obtained in either a multi-group or 'individuals' design is highly suspect.

Conclusion

The mathematical assumptions upon which the PCA is based do not appear to describe the way that data is generated in a typical ERP study. In spite of these violations, the combination of PCA and varimax rotation appears to be able to capture the form of the underlying wave forms fairly well. However the estimation of the amplitudes of component waves in individual records may be erroneous. Both the logical arguments presented here and the simulation results of Wood and McCarthy indicate that the problem is greatest for estimation of the latter components estimated by PCA, and is accentuated if the underlying component loadings and/or amplitudes are correlated.

For these reasons it is suggested that PCA results should be treated with some skepticism. It is worth noting that much of this critique has focussed on biases in the PCA method rather than on instabilities of solution. Thus showing that PCA results can be repeated is somewhat beside the point. The biases will affect each replication.

This critique should not obscure what is perhaps the key point of Donchin and Heffeley's original proposal. The analysis of ERP records can be regarded as a problem in multivariate analysis. The problem is to find an approach within multivariate analysis that does not depend upon mathematical assumptions that have unreasonable physiological implications. The use of PCA was a reasonable first step in the search for such a procedure, but it has some inadequacies. Other multivariate analysis techniques that may be more satisfactory. In a subsequent paper I hope to report an alternative multivariate technique that can be applied to ERP records without making the strong mathematical assumptions inherent in the use of PCA.

References

Donchin, E. and Heffely, R. (1979). Multivariate analysis of event related potential data: a tutorial review. In D. Otto (ed.) Multidisciplinary perspectives on event related potential research. EPA-600/9-77-043: Washington, D.C. U.S. Government Printing Office.

Ford, J.M. and Pfefferbaum, A. (1980). The utility of brain potentials in determining age-related changes in Central Nervous System and cognitive functioning. In Poon, L.W. (ed.) Aging in the 1980s: Psychological Issues. Washington, D.C.

Wood, C. C. and McCarthy, G. (1983). Principal component analysis of event-related potentials: Simulation studies demonstrate misallocation of variance across components. Electroencephalography and clinical neurophysiology. .

Author's Notes

The research reported here was supported by the United States Air Force, contract # AFOSR-83-0289 to the University of Washington. The views expressed herein are the author's, and do not necessarily reflect policy of the sponsor.

I would like to acknowledge the contribution of Prappan Tianwattanatada in preparing the simulation reported here.

Table 1

Amplitudes on original components	Amplitudes on Components Recovered from PCA		
	I	II	III
I	.774	.049	.069
II	.035	.772	.036
III	.009	.429	.784

Correlations between recovered and original components when the original component amplitudes I and II were correlated across records.

Figure Legends

Figure 1: ERPs plotted on two time axes.

Figure 2: Point plotted on time axes (t=1 and 2) and component axes.

Figure 3: Two overlapping ERPs (top) and their average signal produced by them (bottom).

Figure 4: Illustration of the effect of doing PCA on two distinct groups, treated as one data set. The "best fitting ellipse" for the data as a whole may not be illustrative of the data for the two groups individually.

Figure 1

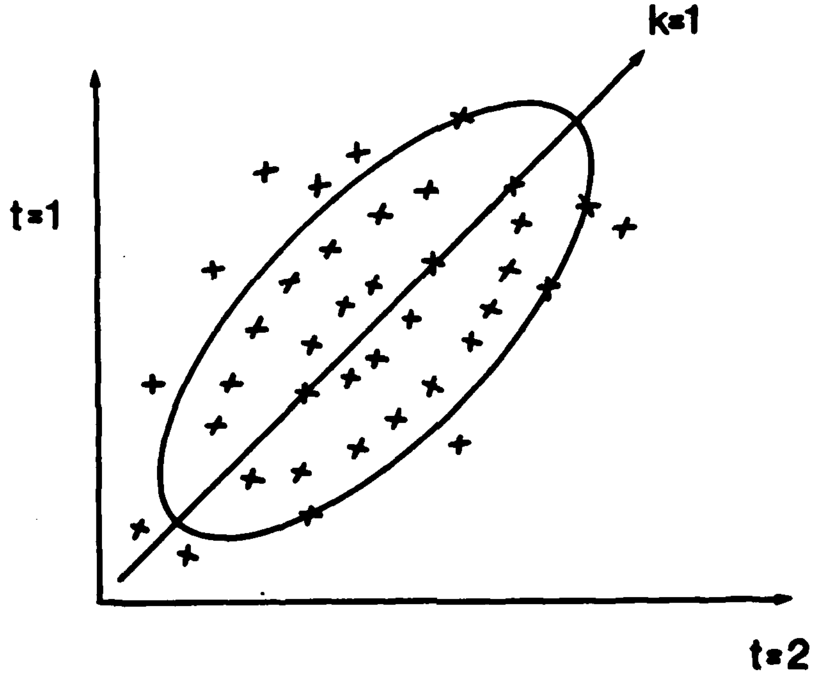
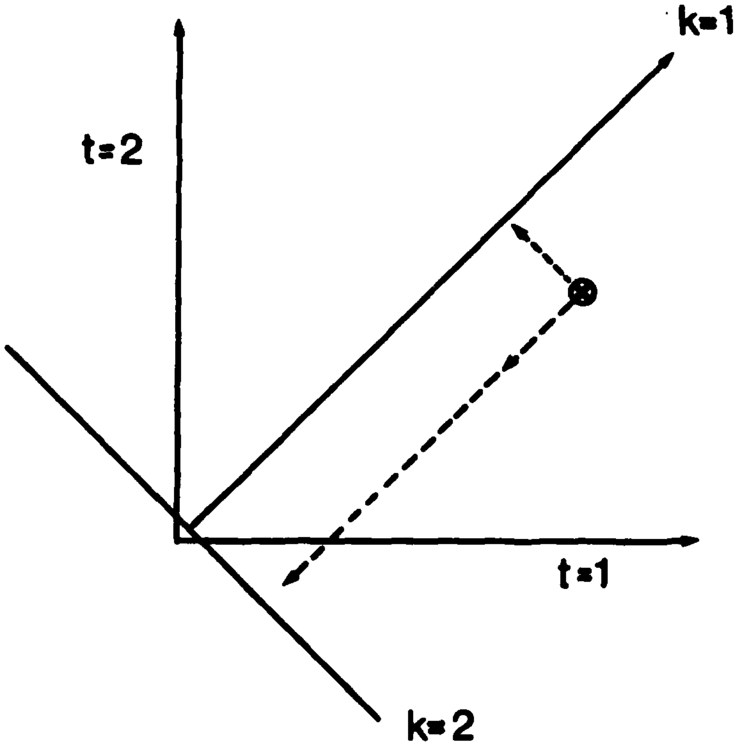


Figure 2



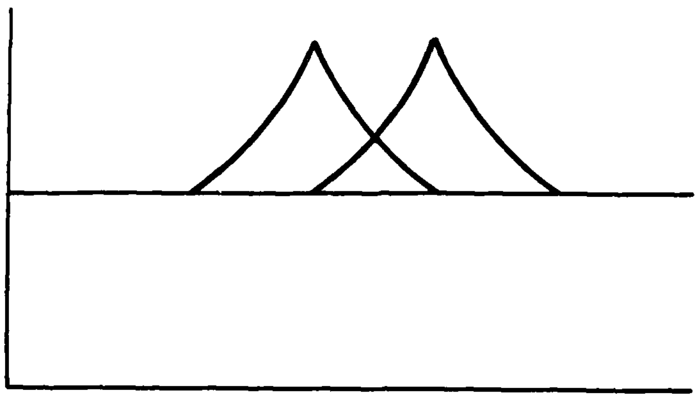


Figure 3a

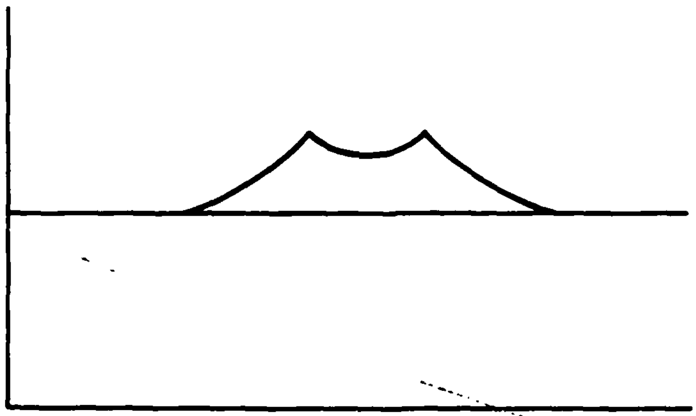
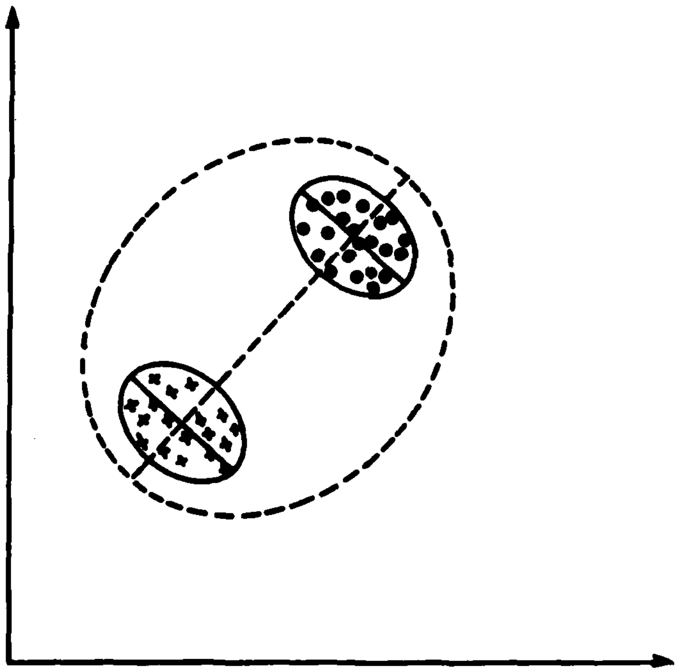


Figure 3b

Figure 4



END

FILMED

9-84

DTIC